THE UNIVERSITY OF ALBERTA

RELIABILITY ESTIMATION WITH RESPECT TO STRENGTH MEASUREMENT

by

(C) TERRY J. BANFIELD

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF ARTS

FACULTY OF PHYSICAL EDUCATION

EDMONTON, ALBERTA

JULY, 1967

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled

RELIABILITY ESTIMATION WITH RESPECT TO STRENGTH MEASUREMENT

Submitted by TERRY JOHN BANFIELD in partial fulfilment of the requirements for the degree of MASTER OF ARTS.

ABSTRACT

The purpose of this study was to establish the reliability of six cable tension measures and two dynamometric measures when utilizing the Hettinger Strength Chair (modified by Howell). A single test group of thirty-two university freshmen of age eighteen years was studied. Random sampling techniques were employed in order to assure as representative a sample as possible. Basic equipment used in this study included a strength machine, cable tensiometry instruments and a Smedley Adjustable Grip Dynamometer. The strength machine was designed to provide body stability and an objective reduplication of test posture. Test items included grip strength, elbow flexion and elbow extension strength and leg extension strength. Subjects were tested on four separate occasions, with three trials given for each test item.

Reliability coefficients of the eight isometric strength measures were found to range from moderately high (.74) to high (.98). A random administration of test items when compared to a standard administration did not result in any significant changes in reliability, and produced only small increases in mean strength scores. The superiority of the right body side over the left was evident in the strength scores. A subsidiary problem was the determination of inter-individual differences and intra-individual differences in each one of the strength items and the effect on these differences over a test-retest period. Both variances tended to remain fairly constant. A second subsidiary problem was whether

the use of best scores rather than average scores resulted in increases in reliability. This failed to materialize. In addition, a comparison of the test-retest method of computing reliability with an analysis of variance method showed no significant differences.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	STATEMENT OF THE PROBLEM	1
	Introduction	1
	The Problem	4
	Subsidiary Problems	4
	Hypotheses	4
	Limitations	5
	Delimitations	5
	Definitions	6
II.	REVIEW OF RELATED LITERATURE	7
	Reliabilities of Reported Strength Tests	7
	The Reliability Coefficient	11
	Factors Influencing Reliability	13
	Correlation of Scores	15
	Motivation and Test Scores	18
	The Number of Trials	20
	Time Between Trials	21
	Joint Angles	22
	Summary	24
III.	METHODS AND PROCEDURES	26
	Selection of Subjects	26
	Test Period	26
	Equipment	26

CHAPTER	PAGE
Description of Strength Tests	28
Experimental Technique	30
Equipment Calibration	31
Statistical Procedures	31
IV. RESULTS AND DISCUSSION	33
Descriptive Statistics	33
Test Reliability	33
Discussion of Strength Scores	41
Comparison of Results to Those of Other Studies	43
V. SUMMARY AND CONCLUSIONS	46
Summary	46
Conclusions	48
Recommendations	49
REFERENCES	51
APPENDIX	56

LIST OF TABLES

TABLE PAGE

I. Coefficients of Reliability for Active Strength Tests . . . 8

II. Test-Retest Reliability Coefficients 11

III. Reliability Coefficients of Four Isometric Tests 15

IV. Summary of Experimental Design 31

V. Summary of Descriptive Statistics 33

VI. Reliability Coefficients--Standardized Design 34

VII. Reliability Coefficients--Randomized Design 36

VIII. Comparison of Reliability as Related to Order of Test Item Presentation 37

IX. Inter- and Intra-Individual Variances--Standardized Design . 38

X. Reliability of Standardized Test Items Computed by the Variance Ratio Method 39

XI. Reliability of Six Standardized Order Test Items Computed by the Variance Ratio--Measurement Error Removed 40

XII. Comparison of Test-Retest Versus Variance Ratio Methods of Computing Reliability--Standardized Design 41

XIII. Mean Strength Scores--Standardized Design 42

XIV. Mean Strength Scores--Randomized Design 42

CHAPTER I

STATEMENT OF THE PROBLEM

Introduction

It is important from the standpoint of human development to know what basic level of strength is required to effectively and efficiently perform daily tasks and activities. At the present, information on the nature of this level at various age groups is indecisive, although one or two recently completed normative studies indicate the possibility of extensive age and sex differences. This investigation in combination and in comparison with others will add to the available knowledge.

Cable tensiometry in the measurement of muscular strength has been used extensively for the past fifteen years. This technique involves the maximum tension a muscle group can apply to a light cable. The main advantage of this method is that body and joint angle can be manipulated to produce the most effective position for the application of optimum strength. Objective recordings of muscular strength are registered on the tensiometer dial. This method is restricted to measurement of static muscle strength.

Reliability coefficients for the test items under investigation in this study have been established in many research articles (4,30,38, 39,40,41,44,45) and these, in general, range from .65 to .99. However, the reliability of using cable tensiometry with a strength table to measure muscular strength, has been questioned by Morris (38) at the University

of Washington and by Elkins (11) at the Mayo Clinic. Criticisms from these investigators are predicated on three basic weaknesses: the difficulty of preventing shoulder rotation, stabilization of the strong individuals and duplication of body position in the retest situation.

The Hettinger strength chair is an apparatus designed to take full advantage of the principles of Cable Tensiometry. Originally designed by Hettinger in Germany and rebuilt by Howell (26) in Canada, this chair lends itself to greater measurement over a wide geographical area because of its compactness and the non-necessity of permanent fixtures usually required by similar apparatus. In addition to this practicality, the chair has been designed to eliminate the aforementioned weaknesses: shoulder rotation, stabilization and reduplication of body position. These latter claims are to be investigated.

This study is naturally concerned with reliability in a broader sense. That is, an extensive normative survey has already been completed on Edmonton school children using the Hettinger Strength Chair, and similar additional investigations are planned across Canada. Because of this intended scope of measurement, it is of paramount importance to have reliability information on the apparatus.

Approaches to estimating reliability define error in slightly different ways and therefore, when different procedures are applied to the same test, each one produces slightly different results (19). It is thus important when evaluating test reliability to be aware of which procedures yield the higher and which the lower estimates of reliability. For example, any fluctuation in score from one time to another is called

error by the test-retest procedure. Here, error is defined as anything which leads a person to get a different score on one testing than he obtained on another testing (19). The test-retest method, then, may either overestimate or underestimate the true reliability of the test. Where several distinguishable sources of measurement error exist, the components-of-variance approach permits an evaluation of the relative importance of each (12).

It is frequently the common and accepted practice by many investigators in the field of strength testing to select, as a subject's strength score, his best performance in a series of two or more trials. This possibly evolved as being the most convenient and the least time-consuming. However, coefficients of reliability have been found by some researchers (1,22,35,48) to change if the researcher chose to correlate individual best scores rather than average scores. Other investigators (20,29,34) are in disagreement. Further, if the data is more or less variable within the group, and also within individuals, the correlation coefficients may again change.

It is evident from the literature that many persons who have used the test-retest analysis to determine test reliability have neglected to give proper consideration to the question of score correlation. Further study is required into the correlation of scores in general, as the test-retest method is widely employed in physical education research.

One important complication in strength research is the problem of order in administering the composite items in a test battery. A standardized order can result in subject fatigue and learning. An important

consideration is to determine if a change occurs in reliability when a standardized test order is compared with a randomized test order.

Therefore, with respect to the foregoing discussion, the major problem and subproblems of this study were developed.

The Problem

This study attempts to determine the reliability of six cable tension measures and two dynamometric measures, when utilizing the Hettinger Strength Chair.

Subsidiary Problems

Subsidiary to the main purpose of this study are several other problems:

1. To determine if a standardized testing order or a randomized testing order causes a change in reliability,

2. to determine if any change occurs in the size of the reliability coefficient when this coefficient is calculated by using the test-retest method or analysis of variance method, and

3. to determine the extent to which reliability coefficients are raised or lowered by correlating the best score with the best score and the average score with the average score.

Hypotheses

1. The null hypothesis for subproblem one asserts that no significant difference exists between reliability coefficients determined from randomized and standardized testing orders.

2. The null hypothesis for subproblem two asserts that no significant difference exists between reliability coefficients determined by correlating best scores and those of average scores.

3. The null hypothesis for subproblem three asserts that no significant difference exists between reliability coefficients calculated by the two different methods.

Limitations

1. Three trials were administered to each subject on a test-retest basis. The retest was given the following day at approximately the same time that the previous day's trials were administered.

2. No control was placed upon the subject's activity on the day of the test except that he was to refrain from strenuous physical exercise for at least one hour prior to the test.

3. A further limiting factor involved in this study was the accuracy of joint angle measurements.

Delimitations

1. This investigation was delimited to thirty-two freshmen registered in the University of Alberta physical education service program, September, 1966.

2. The subjects tested were of age 216 to 224 months.

3. The following tests were used in the assessment of strength:

- 3.1 right grip strength
- 3.2 left grip strength
- 3.3 right arm extension
- 3.4 left arm extension
- 3.5 right arm flexion
- 3.6 left arm flexion
- 3.7 right leg extension
- 3.8 left leg extension

4. The amount of effort exerted by each subject during the testing sessions was accepted to be his maximum.

Definitions

Muscular Strength: the ability of an individual to exert a single explosive force against an object (26).

Obtained Score: the numerical value which actually occurs when a subject takes a test.

True Score: the mean of a hypothetical infinite series of measurements on that subject, each of the measurements being independent of the others and all being taken under the same conditions (12). The true score is represented by inter-individual variance.

Variable Errors: errors that differ from person to person during any one testing and which vary from time to time for a given person, measured twice by the same instrument (19).

Intra-Individual Variance: the variance attributable to biological variation in the functional status of the individual.

Inter-Individual Variance: the variance attributable to true differences between individuals.

Statistical Notations: total variance = σ^2_x

inter-individual variance = σ^2_t

intra-individual variance = σ^2_i

CHAPTER II

REVIEW OF RELATED LITERATURE

A review of the literature related to strength tests indicates the limited availability of objective techniques in this area prior to 1946. The tests proposed by early authors relied principally upon motion against gravity with various degrees of resistance applied by the examiner. Clarke (5), over a period of time, developed tests for measuring the strength of thirty-eight muscle groups using a tensiometer. In the course of this work, apparatus and objective techniques were devised for measuring the strength of muscles activating joint movements in the body. However, Clarke's techniques have been found to involve an element of shoulder rotation, difficulty in stabilizing strong subjects and the problem of duplicating body position in a retest situation (11,38). The Hettinger strength apparatus was designed to utilize cable tension techniques and eliminate the problems of rotation, stabilization and reduplication of body position.

Reliabilities of Reported Strength Tests

In 1925, Rogers (7), initially obtained the following reliability coefficients for two Physical Fitness Index test items: right grip, .92, left grip, .90. These coefficients were obtained from two tests given four months apart.

Rarick, *et al*. (45), in a study of active and breaking strength measurements of the knee and elbow extensors and flexors, utilized

Clarke's (8) cable tension techniques. Coefficients of reliability were computed by correlating the socres from trial one with those of trial two taken on the same day. The results of their investigation are presented in Table I.

TABLE I

COEFFICIENTS OF RELIABILITY FOR ACTIVE STRENGTH TESTS

Age	Sex	Elbow Extensor	Elbow Flexor	Knee Extensor	Knee Flexor
7 Yrs.	Girls	.95	.98	.90	.96
7 Yrs.	Boys	.96	.91	.93	.95
10 Yrs.	Girls	.93	.65	.95	.89
10 Yrs.	Boys	.97	.95	.96	.96

Henry and Smith (21) tested thirty male subjects using the Smedley dynamometer and obtained reliability coefficients of .820 in the dominant hand and .768 in the non-dominant hand, both correlations being for single trials. When reliability coefficients were calculated for the average of two single trials with each hand the following results were obtained: dominant hand, .931, non-dominant hand, .861.

Further investigation by Henry (23) with a Smedley hand dynamometer indicated test-retest reliability coefficients of grip strength for forty-one male and thirty-three female college students. The tests and retests were separated by approximately one week. Removal of measurement error increased the reliability from .753 to .782 for the men and for the women the reliability increased from .876 to .897.

Nelson and Lambert (40) studied the elbow-flexion strength of

nineteen male subjects by the cable tension method. The mean of three trials on each day was used to represent the strength score for the three days of testing. Coefficients of reliability obtained were .90 (the mean of three trials on day one versus mean of three trials on day two) and .92 (the mean of three trials on day two versus the mean of three trials on day three).

Cousins (10) administered a test of grip strength to a randomized sample of forty subjects and obtained a reliability coefficient of .85. This test consisted of two trials, with a time interval of two days before the retest was given.

Bowers (2) tested the hand grip strength of one hundred volunteer subjects on three different hand dynamometers--the Narragansett hand spring dynamometer, the Stoelting adjustable spring type dynamometer and the cable tensiometer. The ages of the subjects ranged from eighteen to twenty-four years. Each subject was given two grip strength trials on the same dynamometer, with a five-minute rest between each trial. A uniform adjustment of the cable tensiometer and Stoelting adjustable dynamometer according to each subject's palm length was used throughout this investigation. Reliability coefficients of .94 for the cable tensiometer, .91 for the Stoelting dynamometer and .89 for the hand spring dynamometer were found, when the subjects' strength scores in trial one were correlated with trial two.

Campney and Wehr (4) employed the test-retest method of calculating reliability to relate test-retest scores at different angles of pull for shoulder flexion and knee extension. Forty-two male and female subjects

were retested at two week intervals. Reliability coefficients for knee extension ranging from .997 to .996 were obtained for the range 80° to 130° of motion.

Nelson and Fahrney (41), testing elbow flexion at 130°, reported reliability coefficients of .96 and .91 for two independent groups of twenty-three and thirty-one subjects, respectively. The test-retest reliability coefficient of .96 was calculated from the best of two trials on two successive days, while the mean of two daily trials on two days was used to compute the reliability coefficient of .91.

Eight measures of isometric strength were tested by Rarick and Oyster (44) using cable tension methods. Forty-eight second grade boys were used as subjects and strength measurements included knee extension and elbow flexion. Three trials were recorded on each of the eight strength measures; twenty-four correlations were computed with the reliabilities, on a test-retest basis, ranging from .68 to .93.

Morris (38) determined reliability coefficients for twelve cable tension tests obtained by testing college women. Scores from trial one were correlated with those of trial two by using the Pearson Product Moment method. Only one trial of each strength test was given to each subject. Coefficients of reliability obtained were: right elbow flexion, .93, right elbow extension, .90, and right knee extension, .95.

Nelson (39) measured static elbow flexion by cable tension procedures. Strength was tested on two successive days, using three trials each day separated by a ten-second rest interval. A reliability coefficient of .96 was found between the respective high scores.

Lucas (33), using the Hettinger strength apparatus in the course of investigating the influence of age and sex on the strength of Edmonton public school children, established the following test-retest reliability coefficients.

TABLE II

TEST-RETEST RELIABILITY COEFFICIENTS

	Grip-Right	Grip-Left	Elbow Flexion	Elbow Extension	Knee Extension
	0.91	0.95	0.96	0.84	0.90
N^x=	23	23	21	21	21

^{x}N contains both male and female subjects.

The Reliability Coefficient

The reliability coefficient is sometimes thought of as an indication of the extent to which a test contains variable errors (19). Henry (24), however, states it is also a measure of the ratio of individual differences to total variation in test scores. Feldt and McKee (12) agree in essence when they define reliability, as the ratio of the variance of true scores to the variance of the obtained scores. Modern statistical texts recognize that the basic concept of the reliability coefficient is derived from the expression $\frac{\sigma^2_t}{\sigma^2_x}$. This relationship, however, can never be computed directly, since the true scores for a sample of examinees are unknown. Nevertheless, the importance of this definition cannot be minimized, for all reliability formulas yield estimates of the value of such a ratio. It should not be inferred that

all reliability formulas represent estimates of the same theoretical ratio. "For before any estimate may be made of the variance ratio, the investigator must identify which factors of the many that influence the obtained score, are to be counted as contributing to the error variance and which to the true-score variance" (12:281).

Feldt and McKee (12:281) further state that:

> Adoption of the test-retest method automatically results in the assignment of certain factors to the error component of each subject's score. One of the most important of these factors is the above or below average performance on a given day. Since this effect is conceived to be constant for any subject on any single day but variable for any subject from one day to another, the test-retest correlation is lowered to the extent that this factor operates. The test-retest technique while fairly simple to apply does not represent the most efficient technique of estimating the ratio of true variance to obtained variance.
>
> The analysis of variance approach is particularily suited to reliability analyses in physical education because of the rather common occurrence of situations in which several components of error variance may be distinguished. Where several distinguishable sources of measurement error exist, the components-of-variance approach permits an evaluation of the relative importance of each.
>
> The analysis also allows the experimenter to estimate the effect of a greater variety of modifications of the original test than can be estimated from the Spearman-Brown prophecy formula.

Helmstadter (19:63) discusses the test-retest procedure, referring specifically to the test-retest method:

> Any fluctuation in score from one time to another is called error by this procedure. Here, error is defined as anything which leads a person to get a different score on one testing than he obtained on another testing.
>
> Test-retest procedure, then, may either overestimate or underestimate the true reliability of the test. Many changes in score are not actually error but intra-individual variation caused by various factors. (18:65)

Helmstadter (10:74) continues: "Since each of the approaches to

estimating reliability defines error in a slightly different way it is not difficult to imagine when different procedures are applied to the same test each one produces slightly different results."

Henry (23) found that the method of computing test-retest reliability as the ratio of "true score" variance to total variance, underestimates the coefficient when the variability of test and retest scores differs by more than 15 per cent. Henry presented a formula for correcting this attenuation. It would appear that intra-individual variations are much larger than the measurement errors in strength testing; if so, they constitute the chief factor that determines test-retest reliability for strength tests. In an attempt to solve this dilemma, Henry (23,24) has suggested that separating inter-individual and intra-individual differences on the basis of test-retest variances tends to give a better estimate of test-retest reliability.

Factors Influencing Reliability

Reliability can be influenced by such extraneous factors as the time of day, the equipment used, momentary attitude of the subject, conditions in the surrounding area such as heat, light, humidity and lack of specific directions for performing the test.

Garrett (15) recommended that practice and the confidence induced by familiarity with the testing apparatus will almost certainly affect the scores when the test is repeated a second time. Moreover, these transfer effects are likely to be different from person to person. He states (15:338): "If the net effect of transfer is to make for closer agreement between scores achieved on the two givings of the test than

would otherwise be the case, the reliability coefficient will be too high."

Guilford (16) pointed out that the following factors affect the reliability of a test:

1. Reliability is highest when the items of the test all intercorrelate highly.
2. The more nearly equal are the difficulties of the test items, the higher is the test reliability.
3. Reliability increases with an increase in test length.

Weiss and Scott (50) report on an investigation by Elbel which advised that test reliability values were influenced by the types of subjects participating in the investigation. Elbel mentioned that it is easier to obtain a high reliability if the subjects range widely in level of achievement (in this case strength) than when they are more nearly equal.

"The test-retest method estimates less accurately the reliability of a test which is highly susceptible to practice than it estimates the reliability of test scores which involve familiar and well-learned operations, little affected by practice" (15:338).

Reliability is also affected by the time interval between testing periods. Kroll (32) has studied the reliability of right wrist flexor strength in a test-retest situation using twenty male subjects on five trials secured on each of three successive days.

Measurement procedures were repeated three weeks later and again in three months. Varying levels of reliability were obtained under each

of the test conditions separately (.91, .99 and .97, respectively).

Strength has also been found to vary at different intervals during the day. Wright (52) found a marked increase in strength of grip from six A.M. to ten A.M.; in some cases a more gradual increase from ten A.M. to one P.M., and a great decrease at night. It would seem advisable, therefore, from the point of reliability in test-retest on different days, to measure strength at the same time of day.

Correlation of Scores

Jones (29) employed cable tensiometry in testing thirty college students on a test-retest basis to determine the reliability of four standardized isometric strength tests. Reliability coefficients using the mean of three trials yielded coefficients quite similar to those of best scores (Table III).

TABLE III

RELIABILITY COEFFICIENTS OF FOUR ISOMETRIC TESTS

Muscle Groups	Best Scores	Mean Scores
Right Hamstring	.766	.757
Right Quadriceps	.875	.875
Left Hamstring	.920	.918
Left Quadriceps	.839	.850

McGraw and McClenney (34) hypothesized that the reliability of tests involving muscular strength and endurance would be increased by using the better or average of trials on successive days, rather than correlating one trial on successive days. A total of 152 boys were

tested on push ups, pull ups and sit ups. The investigators pointed out that when examining the data, a desirable result is one where the t ratio is small and insignificant and the correlation very high. Although coefficients of reliability were in general more favourable for the better of two trials and the average of two trials than for a single trial, the values of the t ratio were as high or higher. Very little difference was found between the use of the better of two trials and that of the average of two trials.

Henry (22), in 1942, found a significantly higher correlation between a vertical jump test and an athletic ability criterion when he used the individual "averages" of the jump scores rather than the "best" of the jump scores.

Smith and Whitley (48), in a further investigation of this approach stated: "While we assumed that this finding would apply to strength tests, there has been no demonstration that the assumption is correct" (48:248). Four measurements of a lateral adductive arm strength were taken at intervals of two minutes for sixty college men. Using the average of the four trials as a strength score the correlation with a speed criterion was r = .66. Using the best score it was .57. The t ratio of the difference, using z transformation formula with a common variable, was 3.3. Consequently, the use of the average gave a significantly higher correlation. They concluded, "It seems evident that the practice of using the best of several performance scores in preference to using the individual averages does not rest on a sound foundation" (48:249).

Berger and Sweney (1) determined that relationships involving the

two best scores resulted in a significantly higher coefficient than relationships between two average scores, provided the variability of scores within groups is sufficiently high relative to the variability within individual scores. They stated:

> The decision to choose an individual's best score, or average score should be based on the degree of variability between all scores relative to the average amount of variability within subjects. The greater the variability between scores relative to the variability within individual scores favors selecting the best score rather than the average score. The opposite selection may be made when the variability between all scores approaches the average variability within individual scores. (1:369)

Henry opposes the findings of Berger and Sweney:

> There is one possible mechanism that could theoretically tend to increase correlations by using best scores without influencing average score correlations. However it would not be operative under the conditions specified by Berger and Sweney as giving advantage to the best score. The necessary conditions are that there are relatively large-within-individual variances in both tests compared to between-individual variances and that individuals of relatively high variability in one test must also have high variability in the other. (20:9)

McNemar discusses variability:

> The size of r is very much dependent upon the variability of measured values in the correlated sample. The greater the variability, the higher will be the correlation, everything else being equal. (36:145)

Henry and Smith (21) computed best-trial reliability coefficients of .820 and .768 for grip strength. When reliability coefficients were calculated from the average of two trials, the reliability increased to .931 and .861, respectively.

McGraw and Tolbert (35) administered six tests of physical ability to 128 junior high school boys on two separate occasions for the purpose of comparing single, best and average methods of obtaining reliability.

Three trials were given on each of the two administrations separated by one week.

McGraw and Tolbert remarked:

> One often hears the comment that when scores on a test vary considerably from trial to trial for each individual the most desirable method of scoring to use is the sum or average of trials. Coefficients of variation of the differences between the highest and lowest scores among the three trials of each administration were obtained as indications of this type of variability. (35:73)

In the main, the largest coefficients of variation were obtained for single trials and the single trial method yielded the smallest coefficient of correlation in every test. However, the smallest coefficients of variation were obtained for the best of three trials and the largest coefficients of reliability were found for the average of three. "There does not appear to be any marked relationship between size of the coefficients of reliability and the variability of the individual test trials or variability of the differences between high and low trials" (35:78).

Motivation and Test Scores

Motivational techniques such as shouting and knowledge of achievement have been shown by some investigators (3,28,43,46) to affect physical performance.

Hellebrandt and Waterland (17), for example, found that the mere observation of others alone had a measurable influence on ergometer performance. It was also stated by Morehouse and Rasch (46) that isotonic exercises probably produce better results than isometric, both from a psychological and physiological aspect. Subjects in both groups of their investigation expressed a dislike for isometric effort.

Several individuals experienced discomfort in some areas tested isometrically. In view of this negative attitude, there may be some question as to whether subjects voluntarily work as hard under isometric conditions as under comparable isotonic conditions.

The findings of Ikai and Steinhaus (28:13) indicate that: ". . .Maximum physiological strength is greater than our measurements of voluntary isometric contractions would indicate." In one study by these authors (28) twenty-five subjects exerted a maximum isometric pull each minute for a duration of thirty minutes. An operator standing behind the unwarned subject occasionally fired a starter's gun, two, four, six, eight, or ten seconds before the pull was exerted. The subjects were also motivated by a shout while exerting the final pull of the session. The "after shot" performance was distinctly higher than after no "shot." The 7.4 per cent improvement in performance was attributed to the shot. The average of single terminal pulls accompanied by a shout disclosed a 12.2 per cent increase over similar performances unaccompanied by shouts or shots.

Burke (3:41) stated:

> It is generally agreed that motivation can affect any physical performance in either direction; that is, it might stimulate some persons to do better than they normally would, or it might serve to decrease the performance ability of others.

The results of a study by Pierson and Rasch (43) indicate isometric strength scores are greater, when the subject has knowledge of his performance than when he does not.

One study found in the literature disagreed with the above studies. Jones (29) attempted to assess the effects of subtle motivation

upon isometric strength scores. Three groups were similarly tested but each was subjected to one of three different conditions of motivation. Conditions were duplicated two days later on the retest. Neither encouragement by itself or combined with knowledge of achievement had a significant effect on performance of highly motivated, normal subjects.

The Number of Trials

Strength testing involves the problem of how many trials should be given to adequately represent the strength of the muscle groups concerned. Two factors in this regard are the development of fatigue and subject familiarization with the testing apparatus. Many studies reviewed made no attempt to standardize the number of trials given in a particular investigation. Hinojosa and Berger (25) reported that in their investigation of the back lift, each subject was given at least two trials. If the second trial produced a higher score than the first, a third was given. Yuhasz (53) recommended the use of three trials to measure back and leg lift strength, while he advised only two trials are required for each hand to determine grip strength. Rarick, et al. (45) advised the use of one unrecorded practice trial on each strength test item, immediately followed by three recorded trials. No explanations were provided by these investigators to account for their recommendations.

Burke (3), in studying the relationship of age to strength and endurance in gripping, allowed each subject only one trial. This procedure was decided on after fourteen subjects had been run through a preliminary test, whereby each subject was given three trials with each

hand. Of the eighty-four grip strength measures taken, only two records were found with a slightly higher grip strength after the first trial. It is worth noting that Burke allowed each subject to obtain some experience with the grip dynamometer before recording his grip strength trials.

Henry (23) using a Smedley hand dynamometer, determined the test-retest reliability coefficients for forty-one males and thirty-three females of college age. Four trials were administered to each subject. Usually the first trial resulted in the highest reading, the second in 6.1 per cent of the tests, the third in 4.1 per cent and the fourth in 3.4 per cent.

Orban (42) tested thirty-five weight lifters on four dynamometer strength tests. Each subject was permitted to repeat a test as many times as he desired during the testing period in order to get his best score. "Rarely, however, were the first two attempts surpassed, and then not substantially" (42:12).

Time Between Trials

Few studies in the literature have investigated the effects of varying time intervals between strength trials to determine the minimum rest period needed between trials to offset fatigue and facilitate maximum scores.

Salter (47) found that five maximal voluntary exertions spaced at intervals of one minute produced little or no fatigue. Hellebrandt, et al. (18) determined that a rest period of about sixty seconds between contractions was necessary to avoid fatigue. Henry and Smith (21)

studied simultaneous versus separate bilateral muscular contractions and utilized a three to four minute rest between each measurement of grip strength. Bowers (2) recommended a five minute rest period between trials given to measure grip strength. Again no explanations were given to account for these rest intervals.

Joint Angles

Differences in muscle strength occur when the joint is tested at varying angles throughout the range of motion, due to an increase or decrease in the mechanical advantage of the limb.

The isometric strength curves of Williams and Stutzman (51) indicated that for elbow flexors the test force is maximal at 90 degrees, and drops off in either direction away from this point. The strength curve for quadriceps extension was characterized by a steep slope. Although the mean curve of the group tested dropped from 120 to 90 degrees, several individuals in the group were found to have higher force readings at the 90 degree position of the joint.

Campney and Wehr (4), in studying the strength differences associated with varying angles of pull for knee extension, found strength to vary as the joint angle changed in magnitude from 80 to 160 degrees in 10 degree increments. The strength differences, which resulted from this variation, were insignificant over more than 60 per cent (80^{o} to 130^{o}) of the normal range of motion for this movement. Joint angles of 140^{o}, 150^{o} and 160^{o} dictated strength values, which were significantly less than the strength observed at any joint angle between 80^{o} and 120^{o}.

Strength curves for the elbow and knee joint movements of sixty-four male subjects were studied by Clarke and Bailey (6). The maximal force of the elbow extensors was obtained at the 40 degree angle, while knee extensor strength showed the best in the range of 100° to 125°.

Clarke (5) after examining the muscle strength exerted throughout the full range of joint motion for elbow flexion, elbow extension and knee extension, selected the following angles: elbow flexion, 115°, elbow extension, 40°, and knee extension, 115°.

In contrast to the findings of Clarke (5) the strength curves of Elkins and associates (11) indicated that muscle power was greatest during elbow flexion at 80° to 90°. Wakim, et al. (49) also found muscle power of the forearm flexors was greatest when the elbow angle was between 80 and 90 degrees.

Elkins, et al. (11) commented that it was difficult to stabilize strong persons and prevent elevation of the shoulder and elbow, when testing elbow extension by the Clarke technique (8). They concluded that the peak of power obtained in earlier studies (5,6,9) at the 40 degree angle, must be due to insufficient stabilization allowing protraction of the shoulder. Elkins, et al. stated: "A similar peak could be obtained in our subjects under those conditions" (11:646).

Salter (47), in an extensive review of measurement methods for muscle and joint function, indicated that no standard posture has been advocated for the use of hand grip dynamometers. "The subject is usually instructed to hold the instrument where he feels that he can exert his greatest force" (47:480). Hunsicker and Greey (27) pointed

out that Erb and Rabinowitsch found subjects could squeeze more, with the elbow extended than with the elbow at a right angle. Fisher and Birmen (14) found that supporting the hand dynamometer with the other hand affected the results considerably.

Summary

The following is a summary of the relevant literature reviewed.

Coefficients of reliability for the test items of this study were also established in numerous articles in the literature (4,30,38,39 40,41,44,45). The coefficients reported in the literature ranged from .65 to .99.

The various approaches to estimating reliability defined error in slightly different ways. A components-of-variance approach was shown to permit an evaluation of the relative importance of each distinguishable source of error. The test-retest method may either overestimate or underestimate the true reliability of a test.

The reliability of a test was reported to be influenced by the equipment used, the time of day, momentary attitude of the subject, conditions in the surrounding area, the time interval between tests and increases in test length.

Coefficients of reliability were found by some researchers (1,22, 35,48) to change, if the researcher chose to correlate individual best scores or average scores. Other investigators (20,29,34) were in disagreement.

Motivational techniques such as shouting and knowledge of achievement were shown by the following investigators (3,28,43,46) to affect

physical performance.

The literature was indecisive regarding both the number of strength trials and the time interval to be permitted between these trials.

Differences in muscle strength were found to occur when the body joints were tested at varying angles throughout the range of motion. These differences were attributed to an increase or decrease in the mechanical advantage of the limb.

CHAPTER III

METHODS AND PROCEDURES

Selection of Subjects

Thirty-two healthy, male subjects of the required age range were randomly selected from the total population of male freshmen registered in the University of Alberta physical education service program. Selection of subjects was carried out by the use of class lists and a table of random numbers. The age range of this random sample was 216 to 224 months.

Test Period

Testing was conducted during the period February 1, 1967 to April 1, 1967, in order to assure reasonable homogeneity of chronological age. The data was collected Monday through Friday of each week, during the regular school day. All of the testing was carried out by the author.

Equipment

Strength testing machine. The strength testing machine, designed by Hettinger and modified by Howell was used for all eight tests of basic strength measured. A vertical pole six feet long was attached to a heavy metal base, three feet square. To the pole was attached a seat which chould be lifted or lowered to accommodate subjects of unequal sizes. Slightly higher on the pole was a shaft to which horizontal arms were attached. To the horizontal arms, two elbow holders were fastened.

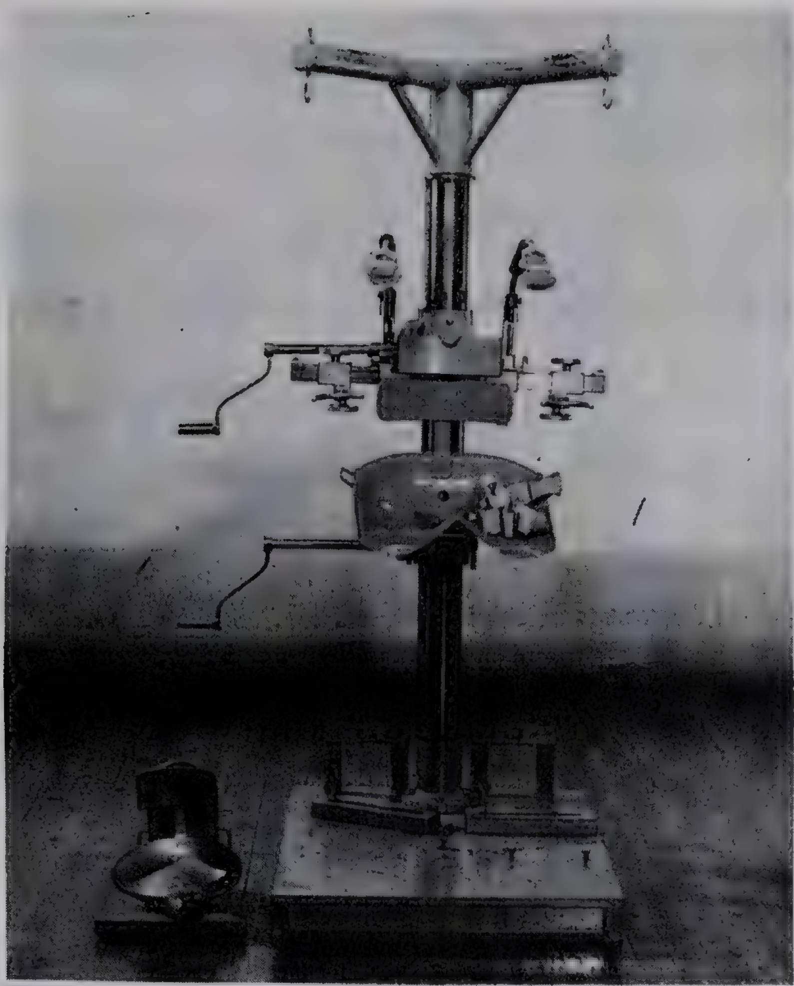

ILLUSTRATION I STRENGTH TESTING MACHINE – FRONT VIEW

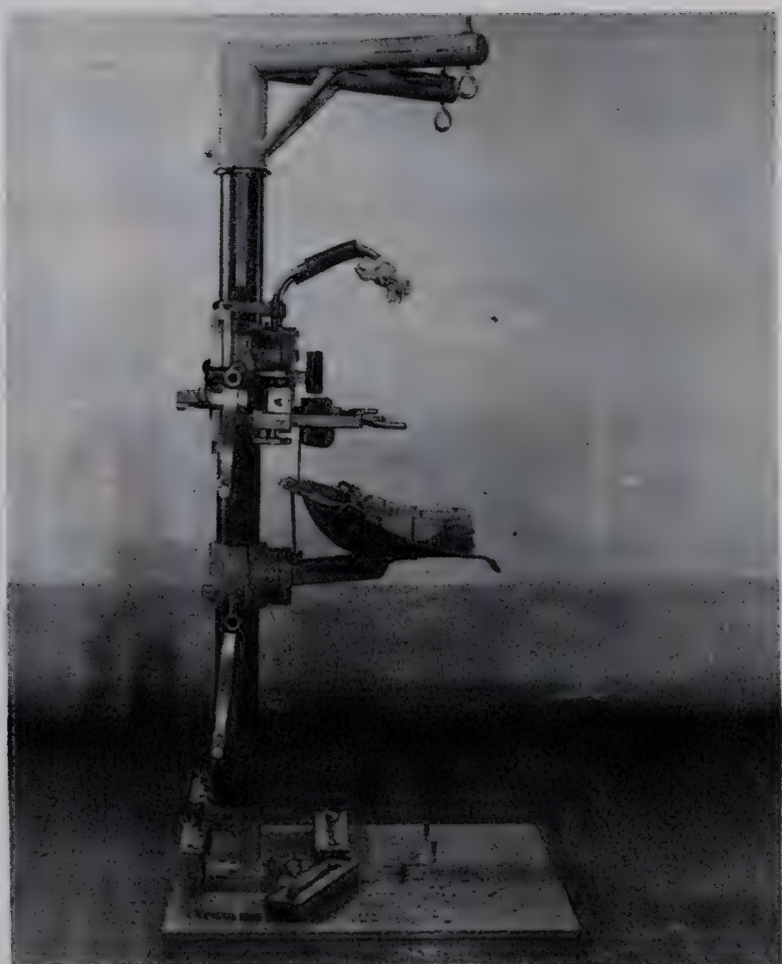

ILLUSTRATION II STRENGTH TESTING MACHINE – SIDE VIEW

The elbow holders stabilized the upper arms both in a vertical line and laterally. They were completely adjustable through the use of a sliding but lockable system. The third attachment to the vertical pole held two shoulder pads, which were adjustable. These adjustable pads stabilized the shoulders to prevent both excessive lifting of the shoulders and forward rotation. A V bar was attached to the top of the vertical pole. This V bar was used as a chain attachment during the elbow extension tests. The base of the apparatus contained a set of adjustable hooks, to which chains could be fastened for the elbow flexion and leg extension tests. The machine was completed by addition of short chains, precision width cables, web belt loops and hooks.

Instrument to record strength of pull. A Pacific Scientific Instrument cable tensiometer was utilized to measure strength of pull for the six cable tension tests. Tensiometer readings were converted directly into pounds by means of a calibration chart.

Goniometer. A goniometer was required to measure the joint angles specified for the various tests. This instrument consisted of a 180° protractor constructed from steel, with two arms, fifteen inches long, attached. One of these arms was stationary extending along the zero line; the other was moveable permitting rotation to the proper angle. A winged nut and bolt placed through the eyelet at the point of rotation of the moveable arm was employed, to maintain set angles of the goniometer.

Description of the Strength Tests

Grip strength test. The strength of grip for both hands was tested by use of a Smedley Adjustable Grip dynamometer. After being seated in the strength machine, the subject was instructed as to the manner of carrying out the squeeze action. The testing arm was flexed at a 90^{o} angle, and the dynamometer was held with the dial facing away from the subject. The grip was adjusted so that the joint between the proximal and middle phalanx fitted over the stirrup of the dynamometer with the hand neither supinated or pronated, but vertically positioned. Turning or rotating of the hand was not allowed. The subject was given six seconds for each contraction. Rest periods of approximately equal duration (under one minute) were allowed, between each maximum exertion, while the instrument reading was taken and recorded. Similar rest intervals between trials were also used for the other test items.

Elbow flexion test. This test was carried out by the use of the strength machine after the shoulder and elbow holders had been adjusted. These adjustments consisted of the subject assuming a comfortable upright sitting position with the shoulders back and evenly balanced. The subject's elbows were positioned against his sides and adjusted forward or backward, so his upper arm was vertical. The hands remained clenched and vertical throughout the test duration. Using a goniometer, the angle at the elbow was adjusted to 120^{o}. A belt loop was then placed around the arm and positioned midway between the wristbone and the olecranon process. A cable and chain was snapped to the loop and

attached to the adjustable hook at the base of the machine. The hook was adjusted to the perpendicular with the lower arm in order to make the angle of pull straight. Using the tensiometer attached to the cable, the subject flexed maximally against the taut cable for six seconds.

Elbow extension test. The overall procedure was similar to the described method of measuring elbow flexion strength. Differences in technique included: adjusting the angle of the elbow to 90 degrees; attaching the cable and chain to the V arm of the strength machine; having the subject extend his lower arm downwards against the taut cable, and emphasizing bending of the lower arm at the elbow rather than pushing. The tendency to push was also opposed by the opposite shoulder pad.

Knee extension test. The subject remained in the strength machine with his hands placed lightly on his legs. A belt loop was placed around the subject's lower leg, midway between the malleolus and the knee bone. The angle at the knee joint was set to 120 degrees. The cable and chain was then fastened to one of a series of hooks located at the back of the strength machine base. Placement of the cable was perpendicular to the lower leg and adjusted laterally so the angle of pull was zero. The tester held an object in line with the proper angle of pull, so the subject's maximal extension was properly aligned.

Experimental Technique

All testing was conducted in one of the physical education research laboratories at the University of Alberta. Subjects were requested to refrain from vigorous physical activity for at least one hour prior to their testing. Each subject selected was required to report to the laboratory on four separate occasions. That is, for one design, he was tested on one day andthen retested at the same time the following day. After a period of two weeks the same subject was again tested using the other design, with a retest given at the same time the following day. Refer to Table IV for a summary of the general design.

Subjects were assigned at random to one of two test designs thereby determining the design in which subjects would first be tested. As there were four different strength tests for the right body side and correspondingly four for the left body side, this resulted in twenty-four possible ways of administering test items on both body sides. A draw was made by each subject to determine his test order for the randomized design. Test items for the standardized design were administered in the following sequence:

1. Right grip strength
2. Left grip strength
3. Right arm extension
4. Left arm extension
5. Right arm flexion
6. Left arm flexion
7. Right leg extension
8. Left leg extension.

For each of the eight test items the subject performed three repeated measures with his right body side followed by three repeated measures with his left body side.

A short explanation of each test and a request for maximum effort was given. Testing was then commenced according to the selected order of administration. Verbal encouragement was used at all times throughout the investigation. The subject was not made aware of his scores during the testing sessions.

TABLE IV

SUMMARY OF EXPERIMENTAL DESIGN

	TEST DAYS Day 1	Day 2	Day 3	Day 4
Subjects N = 16	Randomized Design	Randomized Design	Standardized Design	Standardized Design
Subjects N = 16	Standardized Design	Standardized Design	Randomized Design	Randomized Design
Number of Trials	3 Trials for each test item	3 Trials for each test item	3 Trials for each test item	3 Trials for each test item

<u>Equipment Calibration</u>

A spring-loaded calibration device located at the University of Alberta was used to calibrate the tensiometer employed in this study.

<u>Statistical Procedures</u>

The statistics included the following calculations:

1. Inter-individual and intra-individual variance for all test items on two test days.

2. Test-retest reliability coefficients on the eight items for both test designs.

3. Test-retest reliability coefficients from best scores and average scores for both test designs.

4. Estimation of reliability of test items on standardized design using two different methods.

5. Mean subject scores in pounds on all test items for both the standardized and randomized designs.

A detailed description of the statistical techniques used is found in the Appendix.

CHAPTER IV

RESULTS AND DISCUSSION

Descriptive Statistics

The sample group which was tested was characterized by the following statistics of age, height and weight (Table V). Age was taken as of the first day of testing.

TABLE V

SUMMARY OF DESCRIPTIVE STATISTICS

	Age (Months)	Height (Inches)	Weight (Pounds)
Mean	219.0	70.0	158.6
S.D.	3.0	2.9	23.0
N = 32			

Test Reliability

An acceptable approach for testing the significance of difference between two reliability coefficients from the same group could not be found in the literature reviewed. Furthermore, Garrett (15:242) stated: "Measurement of the significance of difference between two r's obtained from the same sample presents certain complications as r's from the same group are presumably correlated." It was decided, therefore, to use an independent test, Fisher's zr transformation, to test for significance. It is recognized that an independent test is necessarily more powerful

than a correlated test.

The reliabilities of the eight strength measures as calculated by the test-retest method were moderately high (.707) to high (.984) for the standardized design. An analysis of the extent to which reliability coefficients were raised or lowered by correlating best scores and average scores was made. The use of best scores resulted in slightly larger reliability coefficients than the use of average scores for five strength measures. Best and average reliability coefficients for the standardized design are presented in Table VI.

TABLE VI

RELIABILITY COEFFICIENTS--STANDARDIZED DESIGN

Test Item	Best	Average	z Scores
Right Grip	.876	.862	.234
Left Grip	.946	.984	2.538[a]
Right Arm Flexion	.741	.743	.046
Left Arm Flexion	.842	.872	.431
Right Arm Extension	.782	.707	.665
Left Arm Extension	.799	.772	.284
Right Knee Extension	.906	.894	.200
Left Knee Extension	.907	.891	.338

[a]Significant at the .05 level of confidence. A critical z of 1.96 is required for significance.

For only one item, left grip, was a significant difference found

and this was in favor of the average score. However, this result must be viewed with caution. Although the sampling distribution of zr is approximately normal, it exhibits negative skewness for high positive values of r. This makes the interpretation of significance between left grip coefficients .946 and .984 rather difficult. Coefficients for the left body side, which was tested second on all occasions, were slightly higher than those of the right side.

The randomized design resulted in test-retest reliability coefficients of a similar range (.633 to .953) to that of the standardized design. Best score and average score reliability coefficients were compared to determine the magnitude of difference between the two methods. Although the average score method yielded somewhat higher reliabilities than the best score method, for four test items, an analysis of z scores revealed no significant differences between the two score methods. Table VII is a tabulation of best and average reliability coefficients for the randomized design.

It can be seen from Table VII that coefficients for the left body side were slightly higher than those of the right side (for three test items) when the average score method was used. The best score method of computing reliability produced slightly higher coefficients (for three test items) in favor of the right body side.

The standardized test order was compared with the randomized test order to determine if any significant changes occurred in reliability. Test-retest average reliability coefficients were used in this comparison. The random administration of test items did not result in any

TABLE VII

RELIABILITY COEFFICIENTS--RANDOMIZED DESIGN

Test Item	Best	Average	z Scores[a]
Right Grip	.951	.844	1.743
Left Grip	.906	.953	1.388
Right Arm Flexion	.644	.633	.069
Left Arm Flexion	.794	.804	.107
Right Arm Extension	.783	.783	0.000
Left Arm Extension	.771	.824	.561
Right Knee Extension	.915	.919	.123
Left Knee Extension	.880	.878	.026

[a]A critical z of 1.96 is required for significance.

significant changes in reliability when compared to the standard administration. Table VIII shows the results of this comparison of the test orders.

As indicated in Table VIII both orders yielded slightly higher coefficients for four test items. Again, the significant difference between left grip coefficients must be interpreted with caution due to the negative skewness of the zr distribution for high positive values of r.

Inter-individual and intra-individual variances were determined in order to examine the relative size of each over two test days. Alternately, these variances were calculated because they represent two

TABLE VIII

COMPARISON OF RELIABILITY AS RELATED TO ORDER OF TEST ITEM PRESENTATION

Test Item	Standard Order	Random Order	z Scores
Right Grip	.862	.884	.342
Left Grip	.984	.953	2.121[a]
Right Arm Flexion	.743	.633	1.234
Left Arm Flexion	.872	.804	.884
Right Arm Extension	.707	.783	.665
Left Arm Extension	.772	.824	.553
Right Knee Extension	.894	.919	.542
Left Knee Extension	.891	.878	.234

[a]Significant at the .05 level of confidence. A critical z of 1.96 is required for significance.

necessary components in the variance ratio method of computing reliability. Inter-individual variance exhibited little fluctuation except for left arm flexion (an increase from day one) and right arm extension (a decrease from day one). Intra-individual variance was likewise relatively constant over both test days with no common trend indicated. Inter-individual variations were relatively large in comparison to intra-individual variations. Table IX shows inter-individual and intra-individual variances over the two standardized test days.

TABLE IX

INTER- AND INTRA-INDIVIDUAL VARIANCES--STANDARDIZED DESIGN

Test Item	Inter-Individual Variance Day 1	Inter-Individual Variance Day 2	Intra-Individual Variance Day 1	Intra-Individual Variance Day 2
Right Grip	39.32	40.86	4.59	3.70
Left Grip	43.51	42.53	3.69	4.75
Right Arm Flexion	11.88	12.96	1.35	1.79
Left Arm Flexion	12.56	19.11	1.39	0.98
Right Arm Extension	16.26	6.87	1.61	1.66
Left Arm Extension	14.79	17.47	1.57	1.24
Right Knee Extension	43.57	41.22	3.23	3.411
Left Knee Extension	36.69	35.39	3.86	4.05

Table IX demonstrates that strength between subjects varied considerably less for arm flexion and extension than for grip strength and leg extension. Correspondingly, variations within individuals were larger for grip and leg strength measures.

Henry (24) has defined reliability as a measure of the ratio of individual differences to total variation in test scores. Feldt and McKee (12) agree in essence when they define reliability as the ratio of the variance of true scores to the variance of the obtained scores. Statistical texts recognize that the basic concept of the reliability coefficient derives from the expression $\frac{\sigma^2 t}{\sigma^2 x}$. Henry's definition of reliability (24) yields an estimate of such a ratio, where: $\sigma^2 x$

represents total variance (inter-individual variance + intra-individual variance + measurement error variance). Reliability coefficients computed from the variance ratio $\frac{\sigma^2_t}{\sigma^2_x}$ are shown in Table X.

TABLE X

RELIABILITY OF STANDARDIZED TEST ITEMS COMPUTED BY THE VARIANCE RATIO METHOD

Test Item	Reliability Coefficients Day 1	Day 2	Average
Right Grip	.895	.917	.906
Left Grip	.922	.900	.911
Right Arm Flexion	.889	.870	.879
Left Arm Flexion	.892	.945	.918
Right Arm Extension	.903	.793	.848
Left Arm Extension	.896	.927	.911
Right Knee Extension	.928	.920	.924
Left Knee Extension	.901	.894	.897

Coefficients were determined for each of two test days and then averaged for the purpose of later comparisons with the test-retest method. Reliability coefficients, with the exception of left arm flexion and right arm extension, exhibited only small fluctuations from day one to day two. No common trend was evident.

It is evident from the foregoing discussion concerning Henry's definition of reliability (24), that removal of measurement error from total test variance would increase the size of r. Measurement error

variance of the tensiometer was calculated to be .13, and when removed from total test variance had the effect of increasing reliability as shown in Table XI. Measurement error was not determined for the Smedley Grip Dynamometer.

TABLE XI

RELIABILITY OF SIX STANDARDIZED ORDER TEST ITEMS COMPUTED BY THE VARIANCE RATIO--MEASUREMENT ERROR REMOVED

Test Item	Reliability Coefficient Day 1	Day 2	Average
Right Arm Flexion	.896	.878	.887
Left Arm Flexion	.900	.951	.925
Right Arm Extension	.909	.805	.857
Left Arm Extension	.904	.933	.918
Right Knee Extension	.930	.923	.926
Left Knee Extension	.904	.897	.900

The test-retest method and the variance ratio method of estimating reliability were compared. Pearson r average coefficients and variance ratio average coefficients were tested for significance of difference as demonstrated in Table XII.

The variance ratio approach yielded higher but not significantly different estimates of reliability for all measures. Left grip strength was the exception, where a significant difference was found in favor of the test-retest method.

TABLE XII

COMPARISON OF TEST-RETEST VERSUS VARIANCE RATIO METHODS OF COMPUTING RELIABILITY--STANDARDIZED DESIGN

Test Items	Test-Retest Coefficients	Variance Ratio Coefficients	z Scores
Right Grip	.862	.906	.782
Left Grip	.984	.911	3.280[a]
Right Arm Flexion	.743	.879	1.578
Left Arm Flexion	.872	.918	.896
Right Arm Extension	.707	.871	1.140
Left Arm Extension	.772	.911	1.95
Right Knee Extension	.894	.924	.662
Left Knee Extension	.891	.897	.114

[a]Significant at the .05 level of confidence. A critical z of 1.96 is required for significance.

Discussion of Strength Scores

The tensiometer scores were converted to pounds by means of a conversion chart provided by the manufacturer. The Smedley dynamometer scores were calibrated in kilograms, therefore it was necessary to multiply all dynamometer readings by 2.2 for conversion to pounds. Mean strength scores on all strength measures for both test designs were then computed. Tables XIII and XIV represent these mean strength scores obtained by the sample group on the standardized and randomized designs, respectively.

TABLE XIII

MEAN STRENGTH SCORES[a]--STANDARDIZED DESIGN

Test Item	Mean Strength Scores Day 1	Day 2	Average
Right Grip	99.5	99.0	99.25
Left Grip	91.0	89.5	90.25
Right Arm Flexion	82.5	85.0	83.75
Left Arm Flexion	76.5	77.0	77.75
Right Arm Extension	65.5	68.0	66.75
Left Arm Extension	64.0	65.5	64.75
Right Knee Extension	147.0	149.0	148.0
Left Knee Extension	138.5	141.0	139.75

[a]Units in pounds.

TABLE XIV

MEAN STRENGTH SCORES[a]--RANDOMIZED DESIGN

Test Item	Mean Strength Scores Day 1	Day 2	Average
Right Grip	99.5	99.0	99.25
Left Grip	90.5	89.0	89.75
Right Arm Flexion	82.0	87.0	84.50
Left Arm Flexion	78.0	79.0	78.50
Right Arm Extension	66.5	67.5	67.00
Left Arm Extension	66.5	65.0	65.75
Right Knee Extension	153.0	147.0	150.0
Left Knee Extension	141.0	140.5	140.75

[a]Units in pounds.

It can be seen from Tables XIII and XIV that repetition of the test only slightly altered measures of maximal isometric strength. No common trend was apparent with regard to an increase or decrease of mean strength scores on day two. When comparing the standardized design to the randomized design, Tables XIII and XIV indicate that randomization produced only small increases (less than two pounds) in mean strength scores for all test items, with the exception of left grip strength.

Comparison of Results to Those of Other Studies

The following discussion relates the present findings to the results of studies and articles of a similar purpose already carried out. It must be emphasized, however, that although testing techniques were comparable, only one study reviewed made use of the Hettinger strength apparatus. Furthermore, the test group under investigation was comprised of male subjects in a single age group. Frequently, test groups in other studies exhibited a wide age range and involved both male and female subjects. The above points complicated accurate comparisons with other studies.

Insofar as accurate comparisons were possible, reliability coefficients were found to be of the same general range. Grip strength coefficients in the present investigation ranged from .864 to .984. The range indicated in the literature reviewed was .76 to .94.

Arm flexion coefficients under test-retest conditions were lower than coefficients reported in other investigations. However, the variance ratio approach yielded higher reliabilities which were comparable to those of other studies. For comparative purposes the ranges

were .633 to .951 (present study) and .65 to .98 (other studies).

Arm extension strength followed a similar pattern to that of arm flexion. In the present investigation the range was .707 to .933, while in others it was .85 to .97.

Reliability coefficients for leg extension strength ranged from .878 to .930. The range in other reported research was .86 to .99.

No significant advantage was found for the use of best scores over average scores in the computation of reliability. These findings concur with those of the following researchers (20,29,34) while disagreeing with others (1,22,35,48). In particular, the results are in agreement with those of Henry (20) who explains that no advantage is given to the use of best over average when intra-individual variance is relatively small compared to inter-individual variance. In respect to variability of the range of scores both between test items and within test items, the data of this study compared closely to the hypothetical data used by Berger and Sweney (1). However, the present findings are in disagreement with the conclusions drawn by Berger and Sweney (1).

The variance ratio approach to computing reliability resulted in higher but generally non significant estimates of reliability when compared to test-retest. Left grip strength was the exception. In this respect these findings are in agreement with those of Feldt and McKee (12), Helmstadter (19) and Henry (23,24). Helmstadter (19) stated:

> Since each of the approaches to estimating reliability defines error in a slightly different way it is not difficult to imagine when different procedures are applied to the same test each one produces slightly different results.

The test-retest technique does not represent the most efficient approach

to estimating the ratio of true variance to obtained variance, because any fluctuation in score from one time to another is called error by this approach.

CHAPTER V

SUMMARY AND CONCLUSIONS

Summary

It was the primary purpose of this study to establish the reliability of six cable tension measures and two dynamometric measures when utilizing the Hettinger strength chair. Subsidiary problems in the study were the comparison of a standardized test design with a randomized design with respect to change in reliability; the determination of inter-individual variation, intra-individual variation and measurement error for the standardized design; the estimation of reliability by the use of two different methods to determine which method yielded higher or lower estimates of reliability; and an analysis of the extent to which reliability coefficients are raised or lowered by correlating best scores and average scores.

A sampling technique that utilized class lists and a table of random numbers was employed to select thirty-two male subjects of the required age range from the total population of university freshmen.

The strength testing machine was designed to provide reproducible test postures, as well as body stability. In conjunction with the strength chair a tensiometer and loops, chains and cables were also used to test static isometric strength. A Smedley Adjustable Grip Dynamometer was employed for the purpose of measuring strength of grip.

Age, height and weight of the thirty-two subjects were recorded

prior to and during testing. Each subject was measured in eight isometric test items that included: grip strength, arm flexion and extension strength and leg extension strength. Test items for both a standardized and a randomized design were administered according to previously selected orders. All subjects were tested on four separate occasions. Three trials were given for each test item. Subjects were periodically motivated by verbal encouragement without knowledge of their scores.

The reliabilities of the eight strength measures were moderate to high as calculated by the test-retest method (.633 to .984) as well as by the variance ratio method (.848 to .924). In comparison with other studies, the above reliabilities are similar, although arm flexion and extension coefficients were generally lower. It must be understood that although the testing technique was a normal one, only one study reviewed made use of the Hettinger Strength Chair (33). Furthermore, the test group under investigation was comprised of males in a single age group. Frequently test groups in other investigations exhibited wide age ranges and were comprised of both male and female subjects.

A random administration of test items did not result in any significant changes in reliability when compared to the standard administration, though for four test items reliability was slightly higher for the randomized design.

Inter-individual variance was found to be relatively constant over the two standardized test days. Two exceptions were evident: (1) left arm flexion increased from 12.56 to 19.11 tensiometer units,

and (2) right arm extension decreased from 16.26 to 6.87 tensiometer units. Intra-individual variance exhibited little fluctuation for all strength measures over the two days. Inter-individual variance was relatively large compared to intra-individual variance for all items. Measurement error variance was found to be virtually negligible (13).

No significant advantage was found by using best scores over average scores in the computation of reliability.

A variance ratio method of computing reliability yielded higher estimates than the test-retest method with the exception of left grip strength. However, a significant difference between the two techniques was only evident for one test item, left grip strength. Removal of measurement error from total test variance slightly increased the size of the reliability coefficient.

Conclusions

1. The Hettinger strength chair, in combination with cable tensiometry, appears to be a reliable apparatus for measuring the static muscle strength of the test items considered in this study.

2. No significant differences existed between reliability coefficients determined from the randomized design as compared to those determined in the standardized design.

3. No significant differences existed between reliability coefficients estimated by correlating best trials and those estimated by correlating average trials.

4. No significant differences existed between reliability coefficients calculated by test-retest and variance ratio methods.

However, the variance ratio method tended to yield higher estimates of reliability.

5. Removal of measurement error variance from total test variance did not significantly increase test reliability under the conditions of this study.

6. Inter-individual differences were relatively large in comparison with intra-individual variation for the strength measures in this study.

7. Randomization of test items resulted in only small increases in mean strength scores.

8. Generally strength scores for the right body side were higher than for the left body side.

9. Performance on the right body side appears to improve the reliability of repetition on the left side. The reliability in general was higher for the left limb which was tested second.

Recommendations

1. That the strength data obtained from this study be compared with and added to strength data presently available on younger age groups in this province.

2. Further investigation using the same apparatus, to determine the differential motivational effects of no verbal encouragement versus verbal encouragement versus verbal encouragement combined with knowledge of performance results.

3. It is recommended that a study be made to find if performance on one side improves test reliability on the opposite side.

4. It is recommended that a study be made to compare the Clarke Table with the Hettinger Chair, both with respect to mean strength scores and reliability of test items.

REFERENCES

REFERENCES

1. Berger, R. A. and A. B. Sweney. "Variance and Correlation Coefficients," Research Quarterly, XXXVI (1965), pp. 368-69.

2. Bowers, L. E. "Investigation of the Relationships of Hand Size as Measured by Selected Hand Dynamometers," Research Quarterly, XXXII (1961), pp. 308-18.

3. Burke, W. E. "A Study of the Relationship of Age to Strength and Endurance in Gripping," Microcarded Doctor of Philosophy Dissertation, State University of Iowa, 1952.

4. Compney, H. K. and R. W. Wehr. "An Interpretation of the Strength Differences Associated with Varying Angles of Pull," Research Quarterly, XXXVI (1965), pp. 403-12.

5. Clarke, H. H. "Improvement of Objective Strength Tests of Muscle Groups by Cable-Tension Methods," Research Quarterly, XXII (1951), 399-419.

6. _______, and T. L. Bailey. "Strength Curves for Fourteen Joint Movements," Journal of Physical and Mental Rehabilitation, IV (1950), pp. 12-16.

7. _______. Application of Measurement to Health and Physical Education. New York: Prentice-Hall, Inc., 1950, pp. 172-73.

8. _______. A Manual of Cable Tension Strength Tests. Springfield, Massachusetts: 1953.

9. _______, E. C. Elkins, G. M. Martin and K. G. Wakim. "Relationship Between Body Position and the Application of Muscle Power to Movements of the Joints," Archives of Physical Medicine, XXXI (1950), pp. 81-89.

10. Cousins, F. G. "Effect of Trained and Untrained Testers Upon the Administration of Grip Strength Tests," Research Quarterly, XXVI (1955),

11. Elkins, E. C., U. M. Leden, and K. G. Wakim. "Objective Recording of the Strength of Normal Muscles," Archives of Physical Medicine, XXXII (1951), pp. 639-47.

12. Feldt, L. S. and M. E. McKee. "Estimation of the Reliability of Skill Tests," Research Quarterly, XXIX (1958), pp. 279-92.

13. Ferguson, G. A. Statistical Analysis in Psychology and Education. New York: McGraw-Hill Book Company, 1959, pp. 154-55.

14. Fisher, M. B. and J. E. Birmen. "Standardization of a Test of Hand Strength," Journal of Applied Psychology, XXX (1946), p. 380.

15. Garrett, H. E. Statistics in Psychology and Education. New York: Longmans, Green and Company, 1960, pp. 336-39.

16. Guilford, J. P. Fundamental Statistics in Psychology and Education. New York: McGraw-Hill Book Company, 1950.

17. Hellebrandt, F. A. and J. C. Waterland. "Indirect Learning--The Influence of Exercise in the Related Muscle Groups of the Same and Opposite Side," American Journal of Physical Medicine, XLI (1962), pp. 44-55.

18. _______, A. M. Parrish, and S. J. Houtz. "Cross Education: The Influence of Unilateral Exercise on the Contralateral Limb," Archives of Physical Medicine, XXVIII (1947), pp. 76-85.

19. Helmstadter, G. C. Principles of Psychological Measurement. New York: Appleton-Century-Crofts, 1946, pp. 58-83.

20. Henry, F. M. "Best vs. Average." Unpublished paper, University of Berkeley, 1967, pp. 9-10.

21. _______, and L. E. Smith. "Simultaneous vs Separate Bilateral Muscular Contractions in Relation to Neural Overflow Theory and Neuromotor Specificity," Research Quarterly, XXXII (1961), pp. 42-46.

22. _______. "The Practice and Fatigue Effects in the Sargent Test," Research Quarterly, XIII (1942), pp. 16-29.

23. _______. "Influence of Measurement Error and Intra-Individual Variation on the Reliability of Muscle Strength and Vertical Jump Tests," Research Quarterly, XXX (1959), pp. 155-59.

24. _______. "Reliability, Measurement Error and Intra-Individual Difference," Research Quarterly, XXX (1959), pp. 21-24.

25. Hinojosa, R. and R. A. Berger. "Effect of Variations in Hand Grip on Recorded Dynamometer Back Strength," Research Quarterly, XXXVI (1965), pp. 366-67.

26. Howell, M. L. The Development of Strength and Muscular Endurance by Isometric Contractions. Third British Empire and Commonwealth Conference on Physical Education, Perth, 1962, p. 20.

27. Hunsicker, P. and G. Greey. "Studies in Human Strength," Research Quarterly, XXVIII (1957), pp. 109-19.

28. Ikai, M. and H. H. Steinhaus. "Some Factors Modifying the Expression of Human Strength," Journal of Applied Physiology, XVI (1961), pp. 157-63.

29. Jones, R. E. "Reliability of Muscle Strength Testing Under Varying Maturational Conditions," Journal of the American Physical Therapy Association, XLII (1962), pp. 240-43.

30. Kirchner, G. "The Construction of a Bettery of Tests Designed to Measure Strength, Endurance, Power and Speed Among Elementary School-Age Boys." Microcarded Ed. D. thesis, University of Oregon, 1959.

31. Kroll, W. "Reliability of a Selected Measure of Human Stregnth," Research Quarterly, XXXIII (1962), pp. 410-16.

32. _______. "Reliability Variations of Strength in Test-Retest Situations," Research Quarterly, XXXIV (1963), pp. 50-55.

33. Lucas, G. L. "The Influence of Age and Sex on the Strength of Edmonton Public and Separate School Children." Unpublished Master of Arts thesis, University of Alberta, 1966.

34. McGraw, L. W. and B. N. McClenney. "Reliability of Fitness Strength Tests," Research Quarterly, XXXVI (1965), pp. 289-95.

35. McGraw, W. L., and W. J. Tolbert. "A Comparison of the Reliabilities of Methods of Scoring Tests of Physical Ability," Research Quarterly, XXIII (1952), pp. 73-81.

36. McNemar, Q. Psychological Statistics. New York: John Wiley and Sons, 1962.

37. Montoye, H. J. and W. F. Tinkle. "Relationship Between Grip Strength and Achievement in Physical Education Among College Men," Research Quarterly, XXXII (1961), pp. 238-43.

38. Morris, A. "The Effect of a Progressive Program of Exercise on the Strength of College Women." Microcarded Master of Science thesis, University of Washington, 1963.

39. Nelson, R. C. and W. Lambert. "Immediate After Effects of Overload on Resisted and Non-Resisted Speeds of Movement," Research Quarterly, XXXVI (1965), pp. 296-306.

40. _______. "Effect of Overload on Speed of Elbow Flexion and the Associated After Effects," Research Quarterly, XXXVI (1965), pp. 174-82.

41. _______, and R. A. Fahrney. "Relationship Between Strength and Speed of Elbow Flexion," Research Quarterly, XXXVI (1965), pp. 455-63.

42. Orban, W. R. "Dynamometer Strength Tests and Performance of Weight Lifters in International Competition," Journal of Sports Medicine and Physical Fitness, II (1962), pp. 12-16.

43. Pierson, W. R. and P. J. Rasch. "Effect of Knowledge of Result on Isometric Strength Scores," Research Quarterly, XXXV (1964), pp. 313-15.

44. Rarick, L. G. and N. Oyster. "Physical Maturity, Muscular Strength and Motor Performance of Young School-Age Boys," Research Quarterly, XXXV (1964), pp. 523-30.

45. _______, K. Gross, and M. T. Mohns. "Comparison of Two Methods of Measuring Strength of Selected Muscle Groups in Children," Research Quarterly, XXVI (1955), p. 74.

46. Rasch, P. J. and L. E. Morehouse. "Effect of Static and Dynamic Exercises on Muscle Strength and Hypertrophy," Journal of Applied Physiology, XI (1957), pp. 29-34.

47. Salter, N. "Methods of Measurement of Muscle and Joint Function," Journal of Bone and Joint Surgery, XXXVII (1955), pp. 248-49.

48. Smith, L. E. and J. Whitley. "Larger Correlations Obtained by Using Average Rather Than Best," Research Quarterly, XXXIV (1963), pp. 248-49.

49. Wakim, G. M., J. W. Gersteen, E. C. Elkins, and G. M. Martin. "Objective Recording of Muscle Strength," Archives of Physical Medicine, XXXI (1950), pp. 90-100.

50. Weiss, R. A. and G. M. Scott. Research Methods in Health, Physical Education and Recreation. Washington, D.C.: 1964, pp. 213-49.

51. Williams, M. E. and L. Stutzman. "Strength Variation Through the Range of Motion," Physical Therapy Review, XXXIX (1959), pp. 145-52.

52. Wright, V. "Factors Influencing Diurnal Variation of Grip Strength," Research Quarterly, XXX (1959), pp. 110-16.

53. Yuhasz, M. S. "Dynamometer Strength Testing." Unpublished material, Department of P.H.R.E., University of Western Ontario.

APPENDIX

Computation of Test-Retest Reliability Coefficients

Pearson's method in Tables VI and VIII:

$$r = \frac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2 \quad N\Sigma Y^2 - (\Sigma Y)^2}}$$

Where: ΣX = sum of best or average scores on day one

ΣY = sum of best or average scores on day two

N = number of observations.

Computation of Significance of Difference Between Reliability Coefficients

An acceptable approach for testing the significance of difference between two reliability coefficients from the same group could not be found by this investigator. Further, Garrett (11:242) states: "Measurement of the significance of difference between two r's obtained from the same sample presents certain complications as r's from the same group are presumable correlated." It was decided, therefore, to use an independent test, Fisher's zr transformation, to test for significance.

Fisher's zr transformation:

$$z = \frac{Zr1 - Zr2}{\sqrt{1/(N_1-3) + 1/(N_2-3)}}$$

Where:

Zr_1 = transformation of (reliability coefficient) to Zr_1 from Table (E:314).

Zr_2 = as for Zr_1

N = number of observations.

<u>Computation of Inter-Individual Variance for a Single Test Day</u>

$$S^2_{1-2}t = \frac{\Sigma XY}{N} - \frac{\Sigma X}{N} \times \frac{\Sigma Y}{N}$$

Where:

ΣX = sum of trial one

ΣY = sum of trial two

N = number of observations.

$$S^2_{1-3}t = \frac{\Sigma XY}{N} - \frac{\Sigma X}{N} \times \frac{\Sigma Y}{N}$$

Where:

ΣX = sum of trial one

ΣY = sum of trial three

N = number of observations.

$$S^2_{2-3}t = \frac{\Sigma XY}{N} - \frac{\Sigma X}{N} \times \frac{\Sigma Y}{N}$$

Where:

ΣX = sum of trial two

ΣY = sum of trial three

N = number of observations.

Inter-individual variance for a single test day then equals:

$$\frac{S^2_{1-2}t + S^2_{1-3}t + S^2_{2-3}t}{3}$$

Computation of Intra-Individual Variance

DAY 1 RIGHT LEG EXTENSION

Subject Number	Trials 1	2	3		Difference Between Trials 1-2	1-3	2-3
1	39.0	40.5	41.0		+1.5	+2.0	+0.5
2	51.5	49.0	45.5		-2.5	-6.0	-3.5
3	40.0	39.5	36.0		-0.5	-4.0	-3.5
.							
.							
.							
32	27.0	27.5	29.0		+0.5	+2.0	+1.5
N=32				X	-44	-43	-5
				X^2	264	313	180
				MX	1.375	1.343	0.156
				MX^2	8.250	9.781	5.625
				$\sigma^2 i$	3.180	3.989	2.800

For each of Trials 1-2, 1-3, and 2-3,

$$\sigma^2 i = \frac{MX^2 - (MX)^2}{2}$$

Where:

MX^2 = mean of the squared difference

$(MX)^2$ = mean of the difference squared.

Total Intra-Individual Variance for Day 1

$$\sigma^2 i = \frac{3.180 + 3.989 + 2.800}{3}$$

$$= 3.323.$$

<u>Computation of Measurement Error Variance</u>

A spring-loaded calibration device was used to determine the error of the tensiometer employed in this investigation. Twenty recordings were made on the calibration instrument at each of the following settings: 40 pounds, 100 pounds, 160 pounds and 200 pounds. As described by Henry (24), measurement error variance was then computed by the mean square method, for each of these four settings. These four variances were totaled together and divided by four to obtain the average measurement error variance for the tensiometer. The table below illustrates the computation of measurement error variance at the 40 pound setting.

MEASUREMENT ERROR VARIANCE AT 40 POUND SETTING

Correct Reading	Calibrated Reading	Difference From Correct Reading (X)
13	11	-2
13	11	-2
13	10	-3
.		
.		
.		
13	11	-2
N = 20		$X = -47.50$
		$X^2 = +116.25$
		$MX = -2.37$
		$MX^2 = +5.81$

$$\sigma^2 e = \frac{MX^2 - (MX)^2}{2}^{x}$$

$$= .10$$

Where:

$\sigma^2 e$ = measurement error variance

MX^2 = mean of the squared difference

$(MX)^2$ = mean of the difference squared.

[x]This statistic involves the difference between <u>two</u> sets of data, so the variance of a single set is only half as large.

Computation of Reliability by the Variance Ratio Method

$$r = \frac{\sigma^2 t}{\sigma^2 x}$$

Where:

$\sigma^2 t$ = inter-individual variance, and

$\sigma^2 x = \sigma^2 t + \sigma^2 i + \sigma^2 e$.